



Metadata—What Every Attorney Needs to Know

By Donna Payne

Within legal circles, metadata has become a ubiquitous term. It frequently is discussed in journal articles, seminars, and information technology (“IT”) briefings. Even so, many legal professionals do not understand what metadata is and how it works. Simply stated, metadata is information stored within a document that is not evident by just looking at the file. It is an electronic “fingerprint” that automatically adds identifying characteristics, such as the creator or author of the file, the name of individuals who have accessed or edited the file, the location from which the file was accessed, and the amount of time spent editing the file. In addition to data that is automatically added to a document, there is user-introduced metadata, such as tracked changes, versions, hidden text, and embedded objects.

This article focuses primarily on metadata contained in Microsoft® Word (“MS Word”) documents. However, attorneys should be aware that metadata also exists in other commonly used word-processing programs, such as Corel WordPerfect, as well as in other types of programs, such as Microsoft Excel and PowerPoint.

This article is meant to serve as a practical overview of what metadata is and how it works. It examines metadata types and provides information on how legal professionals can eliminate metadata to prevent common pitfalls. The article also discusses recent cases, rulings, and opinions that are related to metadata. It does not, however, attempt to explore the ethical implications related to removal of metadata from documents.¹

How Metadata is Added

Metadata automatically is added to a file when the file is first created and then saved. It also is added when a user opens and edits the document. For instance, when a new document is created, associated “create date” and “author name” metadata are added. Likewise, when a file is printed, the document is tagged with a “printed date.”

Other metadata tags also are added, such as which template was used to create the document and the original author of the template. In addition, there is metadata denoting the full name and path of where the document was stored for the last ten authors of the file. If a template and macro package for document creation or a document management system as a document repository is used, more metadata will be added to the file.

File Properties Dialog Box

The most basic metadata is available through the “File Properties” dialog box. The dialog box can be accessed through the Properties option from the File menu (*see* the sidebar entitled “Accessing the File Properties Dialog Box”). Individuals whose firm uses a document management system may not be able to access this dialog box; however, recipients of the document outside the firm will be able to access it, so it is important to know what information is traveling with the file.

The following is a list of metadata that is stored in the File Properties dialog box:

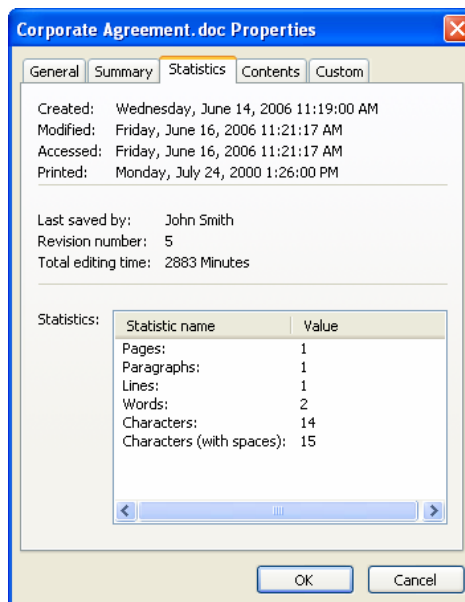
- the version of the software used to create the file
- the full name and file path where the document is located on the computer or network
- file size
- dates of creation, including when last modified and last printed

- summary information, such as the title of the document, subject, author, manager, company name, category, keywords, comments, hyperlinks, and which template was used to create the document
- the name of the person who last saved the document, the number of revisions made, and the total amount of time spent editing the document
- custom properties that may be added to the document, such as the document identification number (client/matter) or a trail of e-mail recipients, along with the subject of the e-mail message to which the document was sent as an attachment if e-mail properties are enabled.

Sidebar 1: Accessing the File Properties Dialog Box

To view metadata in one of your own documents, follow these steps:

1. Open the file.
2. Select the “File” option from the menu, and then choose “Properties” from the drop-down list.
3. Navigate through the General, Summary, Statistics, Contents, and Custom tabs to view properties of the document. Remember that if you use a document management system, you may not have access to this dialog box.



User-Introduced Metadata

The properties of the file represent only a portion of the metadata stored in documents, but exemplify how metadata automatically is added to a file without the knowledge of the end-user. Other metadata is added through the “Field,” “Track Changes,” and “Versions” features.

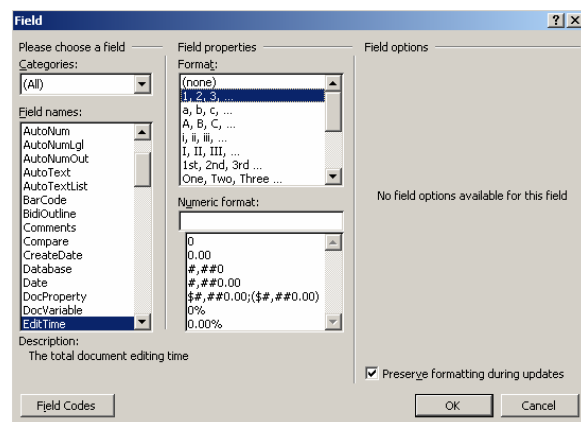
Field Feature

Individuals who do not have access to the File Properties dialog box because their firm uses a document management system may find it helpful to use the Field feature. Like File Properties, the Field feature can be used to view the total amount of time users have spent editing a document. (see the sidebar entitled “Using the Field Feature.”) Many other fields can display metadata or hidden information, as well. For instance, individuals can ascertain the author of the document by inserting the “Author” field and the date the document was created by inserting the “CreateDate” field.

Sidebar 2: Using the Field Feature

Here’s how to find out how long a document has been edited:

1. Open a document that previously was saved.
2. From the Insert menu, choose “Field.”
3. Under Categories, select “All” from the drop-down menu.
4. Under Field names, select “EditTime.”
5. Under Format, select the “1,2,3, ...” option.
6. Click “OK.” The amount of time in minutes a document has been open will appear in the document.



Track Changes Feature

The ability to track all changes made to a document is available in most word-processing programs. When reviewing contracts, it is common practice to hide the changes and show the document in final format or as a final version. However, what happens if an attorney forgets he or she has hidden the tracked changes and sends the document to opposing counsel? The result is accidental disclosure with the potential for an unmitigated disaster. (To learn more about how to use this feature, *see* the accompanying sidebar entitled “Using Track Changes.”)

Sidebar 3: Using Track Changes

Enabling the Track Changes Feature

If you would like MS Word to keep track of revisions you make to a document, you will need to enable the Track Changes feature. Here’s how:

MS Word 97 or 2000: From the Tools menu, choose “Track Changes” and check the option “Track Changes While Editing.”

MS Word 2002 (Office XP) or 2003: From the Tools menu, choose “Track Changes.”

Displaying Other Users’ Changes

If you receive a document that has been worked on with the Track Changes feature enabled, but you don’t see any of the changes or revision marks, you will need to instruct MS Word to display these changes. Once displayed, you can accept or reject them. Here’s how:

MS Word 97 or Word 2000: From the Tools menu, select “Track Changes,” and then “Highlight Changes.” Check the “Highlight Changes on Screen” option, and then select the “OK” button. To accept or reject changes, from the Tools menu, select “Track Changes” and then “Accept or Reject Changes.” Use the “Find” button to navigate through all existing changes in the document, or select the “Accept All” or “Reject All” buttons to affect the changes throughout the document at once.

MS Word 2002 or 2003: From the View menu, choose “Toolbars,” and then select “Reviewing”. The first option on the Reviewing toolbar is the Display for Review drop-down menu. From this menu, select “Final Showing Markup.” The Reviewing toolbar also contains “Accept Change” (look for the blue checkmark) or “Reject Change/Delete Comment” (look for the red “x”) buttons. These buttons can be used to accept or reject individual changes in a document, or to globally accept or reject all the tracked changes in a document.

Versions Feature

Another user-induced producer of metadata is the Versions feature. When this feature is enabled, a subset of the document, all changes, the author of the changes, and the date and time the changes were made are saved in a “micro” version of the document. This data is stored within the document—without any visual indicator or warning that these previous versions reside in the same document. To enable or access Versions, select “Versions” from the File menu. (See Figure 1 to view the result of multiple people working on a document with Versions. Each version is treated as a subdocument within the main document and can be opened with the click of a button.)

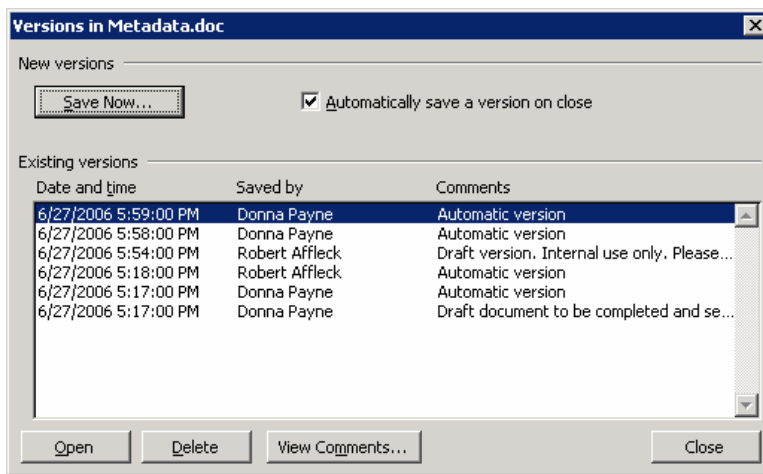


Figure 1. The Versions feature enables users to access previous versions of a document.

Recent High-Profile Exposures

There have been many high-profile documents exposed as having metadata. For example, in 2005, the United Nations released a report on Syria’s suspected involvement in the assassination of Rafik Hariri, Lebanon’s former prime minister. Track changes were enabled in the document and the recipients of the file were able to view names that had been deleted from the document of people said to be involved in the assassination.²

The Pentagon posted a document that revealed that a U.S. soldier accidentally killed an Italian secret service agent in Iraq. The file was improperly redacted and then saved as a PDF file. When the information was copied from the PDF file and pasted into a Word or WordPerfect document, the redacted portions were exposed.

A 51 page document filed with a San Francisco federal court on steroid use in baseball was improperly redacted and then converted to a PDF file format. When the information was copied into Word, eight pages of previously redacted information were revealed.

Metadata discovery was essential in the trial of Merck & Co. and the litigation involving the painkiller Vioxx.³ Through tracked changes that were accidentally left in a document and later discovered by the *New England Journal of Medicine*, it was asserted that the manufacturer knew of the potential heart problem side effects two years before marketing the drug.⁴

What Do the Courts Have To Say?

One of the more detailed opinions on metadata was written about the age discrimination/reduction in force (“RIF”) in *Williams v. Sprint/United Management Company*,⁵ where discovery requests included the native electronic copy of the Excel workbook and its associated metadata. The presiding judge declared that the requested documents should be supplied with the metadata intact, unless both parties agreed that the metadata is not relevant or the producing party requests a protective order. This is significant because it meant that the calculations used to derive the information in the spreadsheet were considered to be metadata relevant to the case.

In the Priceline.com securities class action,⁶ the court ruled that the defendants could use PDF or TIFF file formats for discovery purposes, provided that they also supply searchable

metadata databases and maintain all files in their native software application, with metadata intact, for the duration of the litigation.

These cases and others like them clearly demonstrate that legal professionals must remain vigilant about metadata in documents. Two documents that attempt to define the discoverability of electronic documents and metadata are the Federal Rules of Civil Procedure⁷ and the Sedona Guidelines.⁸ The Sedona Guidelines were published in September 2004 by the Sedona Conference, an organization comprised of judges, lawyers, and scholars.

As metadata and electronic discovery become more prevalent in the courts and in case law, more rules are added to codify the practice of electronic discovery. Recently, the Supreme Court approved e-discovery amendments to the Federal Rules of Civil Procedure, which will go into effect in December 2006.⁹ Most notable with respect to metadata, amendment 26(a)(1)(B) substitutes the words “electronically stored information” for “data compilations” as a category for the required initial disclosures. Rule 26 amendments also cover the exclusions of parties from providing discovery of electronically stored information when the provision of this information is not reasonable because of undue burden or cost; however, the burden remains on the producing party to make the required showing. Rule 34(a) is amended to reference electronically stored information and 34(b) accords parties the right to specify the form or forms of production for electronically stored information sought in discovery.

As more cases involving discovery of electronic documents and metadata are decided by the courts, more rules will undoubtedly be adopted to clarify the role of metadata in discovery.

New York and Florida

Two states, New York and Florida, have already addressed the ethics of searching documents received from opposing counsel or in discovery for metadata, so-called “metadata mining.” Both have expressly declared the practice unethical.

New York State Bar Opinion 782¹⁰ states that “Lawyers must exercise reasonable care to prevent the disclosure of confidences and secrets contained in “metadata” in documents they transmit electronically to opposing counsel or other third parties.”¹¹ The opinion goes on to say, “Lawyer-recipients also have an obligation not to exploit an inadvertent or unauthorized transmission of client confidences or secrets.”¹²

N.Y. State Opinion 749 (2003) concludes that:

The circumstances of the present inquiry present an even more compelling case against surreptitious acquisition and use of confidential or privileged information than that presented by the "inadvertent" or "unauthorized" disclosure decisions. First, to the extent that the other lawyer has "disclosed," it is an unknowing and unwilling, rather than inadvertent or careless, disclosure. In the "inadvertent" and "unauthorized" disclosure decisions, the public policy interest in encouraging more careful conduct had to be balanced against the public policy in favor of confidentiality. No such balance need be struck here because it is a deliberate act by the receiving lawyer, not carelessness on the part of the sending lawyer that would lead to the disclosure of client confidences and secrets.

Nor need we balance the protection of confidentiality against the principles of zealous representation expressed in Canon 7. Our Code carefully circumscribes factual and legal representations a lawyer can make, people a lawyer may contact, and actions a lawyer can take on behalf of a client. Prohibiting the intentional use of computer technology to surreptitiously obtain privileged or otherwise confidential information is entirely consistent with these ethical restraints on uncontrolled advocacy.¹³

The Ethics Committee of the Florida Bar has issued a proposed advisory opinion that directs lawyers to take reasonable steps to safeguard metadata in documents and instructs lawyers who are recipients of documents to not purposefully view metadata that is not intended for the recipient.¹⁴ The proposed rule excludes metadata “that is discoverable under applicable rules or is admissible in a trial or arbitration.”¹⁵

The Problem with These Rules and Opinions

One difficulty with any hard and fast rule on metadata is proliferation of versions and software applications used throughout the legal community. In MS Word, versions of the software process, store, and purge metadata differently. For instance, MS Word 97 stores more author information than later versions. MS Word 2002 and 2003 include options to automatically display residual tracked changes and comments when a file is opened or saved, thereby showing residual tracked changes through no deliberate action to seek out the hidden information. Unless all lawyers are on the same platform and version, the playing field is not level.

System security policies are also available to large firms who use automated installation packages such as the Custom Installation Wizard from Microsoft for configuration and software deployment. These tools only work with Enterprise versions of Microsoft Office and are not available to sole practitioners or firms that purchase the software through original equipment manufacturers. The policy templates also can be configured to control the amount of data that is gathered.

In the case of IT, unfortunately, size does matter. Small firms have fewer tools available to them and often lack the resources to hire a dedicated IT staff tasked with keeping the firm software up-to-date and compliant with technological demands.

How Attorneys Can Protect Themselves

The best way for attorneys to protect themselves is to know what metadata exists in their documents before sharing the documents electronically. For example, knowing whether the “Track Changes” feature has been enabled and whether there are any residual changes in the document is important. Both must be cleared to protect the attorney from accidental disclosure.

Attorneys also should ensure no other versions of the document are stored in the file. In MS Word, this is done by choosing the “File” option from the menu bar and selecting “Versions.”

Finally, view the Properties dialog box to make sure no proprietary information is displayed. It may not be possible to remove all of the built-in file properties; however, individuals should get into the habit of checking this information. “Title,” “author,” and custom properties are added to a document automatically and may contain outdated information or data that should not be disclosed to an outside party.

Microsoft Office XP and 2003 for Windows offers an additional security measure for removing metadata. To access this feature in MS Word, from the Tools menu, select “Options,” and then select the “Security” tab (*see* Figure 2). Next, select the box next to the “Remove Personal Information From File Properties on Save” option. This action is document-specific, which means it needs to be re-selected for every document, unless the user deploys a third-party solution that automatically does this. Being aware of and using these options is a good starting point; however, they fall far short of what is necessary and only deal with a fraction of the total metadata that resides within each file.

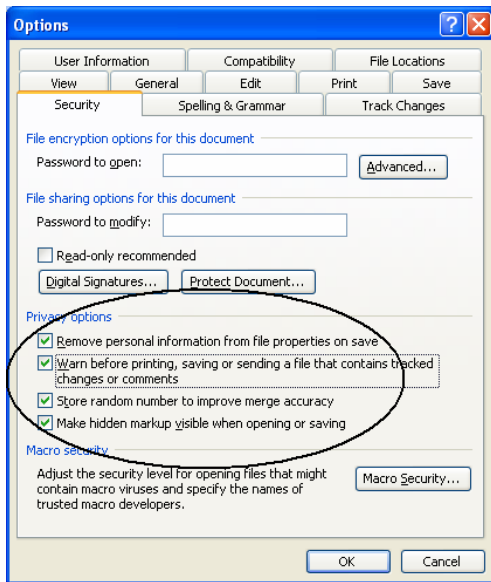


Figure 2. A security feature in Word enables users to remove some of the metadata generated within documents.

Is PDF Safe?

With any electronic file transfer, it is critical to know what is in the file and take precautions to remove any confidential data prior to distribution. Documents that have been properly scrutinized and “sanitized” before being converted into PDF files will not expose redacted information.

The U.S. National Security Agency released a guide to removing metadata from MS Word and PDF documents.¹⁶ This guide explains proper and improper redaction techniques and the preferred method for preparing and distributing sensitive documents.

Redaction software should be used to obfuscate material that is not intended for general viewing. Another effective method for redaction is printing a hard copy of the file, marking it up, and then scanning the document to create the electronic file. It may sound “old-school” to some, but it ensures that information not meant for general viewing is not disclosed. It is imperative for legal professionals to understand applicable court rules before altering any original document.

Some redaction tools create a new document; thus, violating some court requirements that metadata remain with the document and only pre-agreed upon information is redacted.¹⁷

An Adobe[®] Acrobat user can set additional options to protect PDF files from accidental disclosure. In particular, two options can be disabled that control the amount of metadata saved within the document. Select “Adobe PDF” from the menu bar within the word-processing application; choose “Change Conversion Settings” from the drop-down menu; and then uncheck the options “Convert Document Information” and “Attach Source File to Adobe PDF.” (See Figure 3, which shows the Settings tab in Adobe Acrobat 7.0. Prior versions also include these options.)

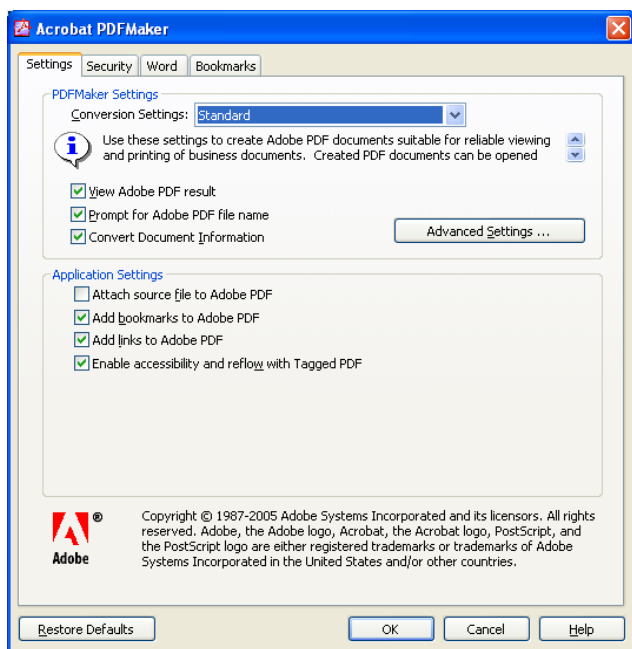


Figure 3. The “Convert Document Information” and “Attach Source File to Adobe PDF” options should be unchecked to prevent information from traveling with the document.

Protection options that prohibit copying and pasting also can be set in Adobe Acrobat to further increase the effectiveness of the software. Just make sure to password-protect the file so that people who have Adobe Acrobat cannot change security settings.

Third-Party Metadata Removal Software

Deploying a third-party metadata removal tool is the attorney's best line of defense in the fight against unwanted metadata. The tools that are available on the market today offer varying levels of protection. Some are much more effective than others. When comparing different products, look closely at what they offer as far as levels of analysis and removal. Check to see whether the tool offers additional protections and features such as PDF conversion, e-mail integration, and customization to meet the needs of the office environment.

Metadata Assistant by Payne Consulting Group was first on the market and is still the most widely used metadata removal product today.¹⁸

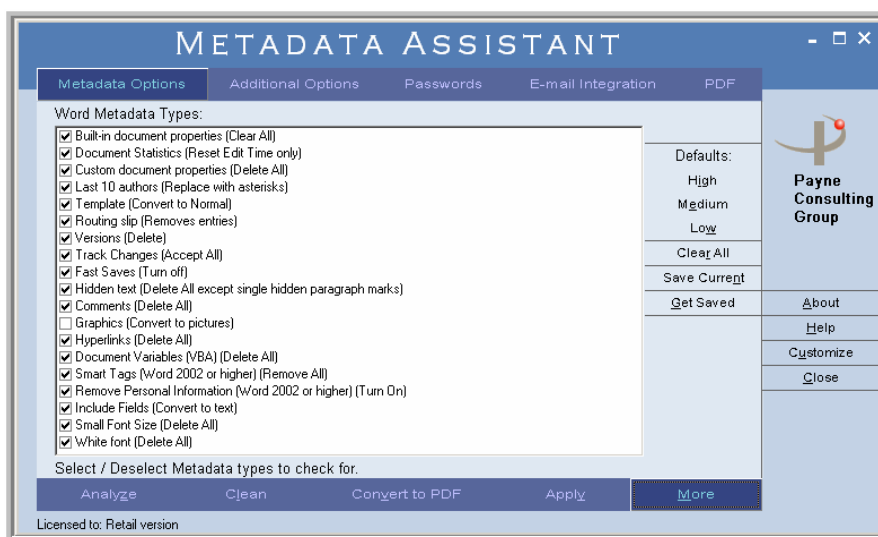


Figure 4. Metadata Assistant by Payne Consulting Group removes residual information from Word, Excel and PowerPoint files and integrates with most e-mail and document management system software. Metadata Assistant also includes a built-in PDF conversion utility.

Conclusion

Metadata often is considered “the smoking gun” in litigation and there appears to be a more concerted effort to obtain it in discovery. It is equally important for sole practitioners, larger law firms, and in-house counsel to be aware of their ethical obligation with respect to the removal of metadata and protection of attorney-client privileged information.

The first step toward accomplishing this is to create internal policy on how future incoming and outgoing documents should be handled. Also, it is productive to evaluate how historical files are introduced into the organization and address retention policy, and to evaluate technology that can be put into place to help secure confidential information and metadata removal and purchase additional software, if applicable. Finally, it is imperative to remain up-to-date on opinions, ethical rules, and cases involving metadata. Obtain a copy of the Sedona Guidelines¹⁹ on working with electronic documents along with the NSA document²⁰ referenced in this article.

Lawyers also must also consider whether there is an obligation to warn clients of the issues associated with metadata. Some very basic questions that every attorney needs to ask include:

1. Do lawyers have a duty to warn clients of the metadata risk?
2. Do lawyers have any kind of duty, in zealously representing the interests of their clients, to look at the metadata in incoming documents?
3. If clients are warned, how do lawyers handle the public relations challenge of explaining why it has taken so long to bring this to light?

Metadata is a component of most, if not all, software that produces electronic documents. It is something that no attorney can afford to ignore.

Donna Payne, Seattle, is president of Payne Consulting Group, a development and training company specializing in law firms, corporate, and government legal departments—(206) 344-8966, donnapayne@payneconsulting.com. Payne is an original member of Microsoft Legal Advisory Counsel, the American Bar Association, the American Society of Journalists and Authors, and the Project Management Institute. Payne Consulting Group is creator of the Metadata Assistant software used by more than 1.7 million people worldwide; as well the other Assistants (Numbering, Forms, Pleadings, Agreement). Payne has authored 12 books on Microsoft Office software.

NOTES

¹ For a more in-depth article on metadata, *see* Zall, “Metadata: Hidden Information in Microsoft Word Documents and Its Ethical Implications,” 33 *The Colorado Lawyer* 53 (Oct. 2004).

² Bone and Blanford, “UN office doctored report on murder of Hariri,” *The Times Online* (Oct. 22, 2005), available <http://www.timesonline.co.uk/article/0,,251-1837848,00.html>.

³ Langreth and Herper, “Merck’s Deleted Data,” *Forbes Magazine*, (Dec. 8, 2005), available at http://www.forbes.com/home/sciencsandmedicine/2005/12/08/merck-vioxx-lawsuits_cx_mh_1208vioxx.html.

⁴ *Id.*

⁵ *Williams v. Sprint/United Management Company*, 230 F.R.D. 640 (D.Kan. 2005).

⁶ *See In re Priceline.com, Inc., Sec Litig.*, No. 3:00CV01884 (DJS), 2005 U.S. Dist. LEXIS 33636 (D.Conn. Dec. 8, 2005).

⁷ *See* Fed. R. Civ. P. 34.

⁸ *Best Practice Guidelines & Commentary for Managing Information and Records in the Electronic Age*, available at http://www.thesedonaconference.org/content/miscFiles/publications_html?grp=wgs110.

⁹ For the text of the Federal Rules of Civil Procedure Amendments, *see* http://www.uscourts.gov/rules/EDiscovery_w_Notes.pdf.

¹⁰ *See* http://www.nysba.org/Content/NavigationMenu/Attorney_Resources/Ethics_Opinions/Committee_on_Professional_Ethics_Opinion_749.htm.

¹¹ *Id.*

¹² *Id.*

¹³ *Id.*

¹⁴ *See* <http://www.floridabar.org/tfb/TFBETOpin.nsf/basic+view/0A1B5E3A86DF495A8525714E005DD6FD?OpenDocument>.

¹⁵ *Id.*

¹⁶ National Security Agency, “Redacting With Confidence: How To Safely Publish Sanitized Reports Converted From Word to PDF,” (Dec. 2, 2006), available at <http://www.nsa.gov/snac/vtechrep/I333-TR-015R-2005.PDF>.

¹⁷ Payne Consulting Group currently offers a free Scramble and Redaction tool (<http://www.payneconsulting.com>) as does Microsoft (<http://www.microsoft.com>). Other paid products are available; however, most do not offer any more substantial benefit over the free utilities.

¹⁸ For more information on Metadata Assistant, see www.payneconsulting.com.

¹⁹ *Supra* note 8.

²⁰ *See* Fed. R. Civ. P. 34.